

O62 Validating and clarifying PROMIS measurement scales using a specialized large language model

Yining Lu¹, Kai Larsen¹, Sen Yan¹, Roland Mueller², Lan Sang¹

¹University of Colorado, Boulder, USA. ²Berlin School of Economics and Law, Berlin, Germany

Objective: Measurement precision in patient-reported outcomes (PROs) remains a critical yet complex challenge due to construct overlap and item ambiguity. This study leverages a specialized large language model (LLM) to validate and elucidate item structures across seven PROMIS item banks: Social Isolation, Physical Function & Mobility, Fatigue, Emotional Distress Anger, Depression, Anxiety, and Alcohol Use. Our objective was to systematically examine internal dimensionality and cross-construct relationships using advanced natural language processing methodologies.

Methods: We employed a fine-tuned LLM tailored for psychometric item bank analysis. Each PROMIS item bank underwent exploratory factor analysis with promax rotation to assess internal consistency, cross-loadings, and dimensional misalignment. Items were evaluated based on standardized loading thresholds for alignment with intended constructs, evidence of unintended cross-loadings, and latent dimensional ambiguity.

Results: The specialized LLM effectively validated the dimensional integrity of the Social Isolation and Emotional Distress (Anger and Anxiety) item banks. Physical Function & Mobility items largely conformed to their expected structure, though notable cross-loadings emerged with Exercise and Health Limitation domains, indicating subtle construct overlap. The Fatigue dimension showed strong coherence, albeit with some item-level overlap involving Social Isolation and Exercise constructs. Alcohol Use presented internal multidimensionality, capturing both dependency and expectancy, but also exhibited substantial overlap between negative consequences and dependency-related items, signaling opportunities for refinement.

Conclusions: The use of a domain-specific LLM enabled a nuanced evaluation of construct validity and dimensional clarity within PROMIS item banks. Findings underscore the potential of AI-driven psychometric methods to enhance the precision of PRO measurement. This approach offers a replicable framework for future validation efforts and scale optimization across diverse health-related constructs.